Journal of Statistical Physics, Vol. 101, Nos. 1/2, 2000

Beyond Storage Capacity in a Single Model Neuron: Continuous Replica Symmetry Breaking

G. Györgyi[1] and P. Reimann[2]

Received December 10, 1999; final March 20, 2000

A single McCulloch–Pitts neuron, that is, the simple perceptron is studied, with focus on the region beyond storage capacity. It is shown that Parisi's hierarchical ansatz for the overlap matrix of the synaptic couplings with so called continuous replica symmetry breaking is a solution, and as we propose it is the exact one, to the equilibrium problem. We describe some of the most salient features of the theory and give results about the low temperature region. In particular, the basics of the Parisi technique and the way to calculate thermodynamical expectation values is explained. We have numerically extremized the replica free energy functional for some parameter settings, and thus obtained the order parameter function, i.e., the probability distribution of overlaps. That enabled us to evaluate the probability density of the local stability parameter. We also performed a simulation and found a local stability density closer to the theoretical curve than previous numerical results were.

KEY WORDS: Neural networks; perceptron; storage capacity; replica method.

1. INTRODUCTION

In his seminal paper Hopfield[1] reformulated Little's model[2] of auto-associative memory in terms of an energy function. By this act, the field of the statistical mechanics of neural networks was plowed and sown in, and proved itself since then remarkably fertile. The network is an interconnected set of McCulloch–Pitts neurons,[3] the latter being perhaps the biologically least realistic model of a nerve cell. In its simplest version the model neuron can be in one of only two states, "firing" or "quiescent," it

[1] Institute for Theoretical Physics, Eötvös University, 1518 Budapest, Pf. 32, Hungary. E-mail: gyorgyi@glu.elte.hu

[2] Theoretische Physik, Universität Augsburg, 86135 Augsburg, Germany. E-mail: reimann@physik.uni-augsburg.de

0022-4715/00/1000-0679$18.00/0 © 2000 Plenum Publishing Corporation

is nonlinear, and admits a large number of connections from other units. Despite the oversimplification in its node cells, the network can exhibit complex behavior and functions as a reasonable model of content addressable memory. From the practical viewpoint, however, artificial neural networks are generically not superior to other methods in the task of storage and associative retrieval.[4, 5]

An important ingredient of this type of models of neural networks is what is called in statistical mechanics quenched disorder. For example, in the Little–Hopfield model the synaptic coupling strengths, made up of random numbers by the Hebb rule,[6, 5] can be held fixed for the duration of the neural dynamical process. This is the basis for the analogy between spin glass models, the archetypical example of a many-body problem with quenched disorder, and neural networks.[6, 7] Methods borrowed from the theory of spin glasses, in particular from the infinite range interaction Sherrington–Kirkpatrick model, yielded a harvest of results on a variety of neural model systems, as well as on other problems whose motivation came from outside of physics but could be formulated as disordered statistical mechanical systems.[5–7]

It is safe to say that the technique inherited from spin glass theory and the most widely used in the statistical mechanical approach to neural networks in equilibrium is the replica method.[6] It was first applied thoroughly for infinite range interaction spin glasses, thus it is especially suited for networks where neurons have a large number of connections. In its simplest version, with so called replica symmetry, it can be straightforwardly adopted to many a neural problem.[5]

Another family of artificial neural networks consists of layered feedforward networks,[8] which accept a number of inputs and for a given set of couplings produce the output as a function of the inputs. Such networks were introduced for the purpose of generalization, i.e., rule extraction, from examples of input–output pairs. Then the *a priori* unknown target rule is to be reconstructed by adaptive changes in the couplings, called training. Since the introduction of the backpropagation algorithm[9] for that purpose, feedforward networks found wide usage.[5, 10, 11]

The statistical physics of neural modeling gained an impetus of lasting effect from the work by Gardner and Derrida[4, 12] on the storage problem of a single neuron, called also simple perceptron in that context. This is the simplest version[13] of feedforward networks. While in the Little–Hopfield statistical mechanical system the quenched variables are the couplings and a microstate is a configuration of neural states, the roles in feedforward networks are reversed. Now the space of synaptic couplings is considered as the configuration space within Boltzmannian thermodynamics and the examples appear as quenched parameters. The error for one example

measures the difference between the actual output of the network and the required output. The errors on all training examples add up to form the Hamiltonian, i.e., the cost function, of the statistical mechanical model. Within the canonical statistical mechanical approach the temperature plays its usual roles. On the one hand, it is the Lagrange multiplier associated with a preset value of the error, on the other hand, it is the amplitude of noise if a gradient-descent-like dynamics of the couplings is used so as to reach optimal configuration. The thermodynamical limit is achieved by admitting a large number of adjustable couplings, but for that it is not necessary to have many neurons. The approach of Gardner and Derrida was successful in what is called equilibrium learning, when a Hamiltonian can be associated with the problem. However, statistical physical methods are proving themselves useful also in studying off-equilibrium learning algorithms.[14]

A central quantity of a feedforward network is its storage capacity, i.e., how many random input–output examples the network can reproduce without error. In terms of the statistical mechanical approach this is in its original formulation a zero temperature problem. The subspace of couplings that reproduce a given set of patterns is called version space, its volume, related to the ground state ($T=0$) entropy, vanishes beyond capacity. Since the statistical mechanical solution of the region below capacity of a single neuron by Gardner and Derrida[4, 12] a number of results have been obtained about storage properties of feedforward networks, see for example refs. 11, 15–18. Nevertheless, if the task is to minimize the number of incorrectly stored examples, beyond capacity the problem has not been solved. Technically this is because below capacity for completely random examples replica symmetry holds, while beyond it no finite replica symmetry breaking scheme yields thermodynamically stable solution.[19]

In the present paper we reconsider the problem of storage of random patterns, technically generalize Parisi's solution of the Sherrington–Kirkpatrick model, and obtain beyond capacity a phase reminiscent to the frustrated ground state of the Sherrington–Kirkpatrick model. That phase continues for $T>0$ into the analog of the low temperature, or Parisi, phase of the spin glass.

Our work is motivated not only by the problem of storing random patterns. Generalization has also been successfully analyzed by statistical mechanical methods, see refs. 10, 20 for reviews. The storage problem below capacity is analogous to equilibrium learning of a learnable task, where the network is compatible with all possible examples, there is no frustration in either systems. For instance, equilibrium generalization properties of the perceptron when the examples are generated by another

one, can be understood for arbitrary number of examples within replica symmetry.[21] However, a given feedforward network may be unable to reproduce a complicated target function on all possible examples.[10] If the target is unlearnable then the network is presumed to get into a frustrated phase if a sufficiently large number of examples are used. Thus the properties of a network beyond capacity are of foremost interest from the viewpoint of rule extraction as well.

The single neuron may be considered as the hydrogen atom of neural problems and studied for its own interest. It is the unit of the Little–Hopfield network, where the symmetry in the couplings has been given up and the couplings of different neurons are considered as independent. As to a feedforward network, even if the whole network operates without error, its units may still be strained beyond their individual capacities.[15] Thus the description of a single neuron beyond its storage capacity is of importance also from the viewpoint of networked neurons. Furthermore, a close analogy exists between the behavior of model neurons beyond capacity and the glassy, frustrated, phase of disordered spin systems.[6, 7, 22, 23, 19, 24] Therefore, the understanding of the way a single neuron works may have ramifications beyond the field of artificial neural networks.

We firstly review in Section 2 the statistical mechanics of storage, recall the basic thermodynamical quantities and the formula for the replica free energy of a single neuron. In Section 3 we summarize the main ingredients of Parisi's approach and obtain the free energy functional that we propose describes the equilibrium problem. This way some background is given to our previous communication,[24] wherein we identified a spin glass phase of Parisi type in the high temperature limit. The recipe for the calculation of expectation values by means of Green functions is explained in Section 4, producing among other the formulas for the local stability distribution and the stationarity conditions. Section 5 is devoted to the scaling for low temperatures, enabling us to put the extremization of the free energy functional on a computer, and in the end a simulation is discussed.

2. THE STORAGE PROBLEM AND ITS REPLICA FREE ENERGY

The model neuron, or perceptron, under consideration is[3, 8]

$$\xi = \mathrm{sign}(h) \tag{2.1a}$$

$$h = N^{-1/2} \sum_{k=1}^{N} J_k S_k \tag{2.1b}$$

where $\boldsymbol{J}$ is the synaptic coupling vector, $\boldsymbol{S}$ the input and ξ the output. Patterns are given as input–output data,

$$\{\boldsymbol{S}^{\mu}, \xi^{\mu}\}_{\mu=1}^{M} \tag{2.2}$$

In the simplest setup the S_k^{μ}s are independently drawn from any distribution with unit variance and zero average, and $\xi^{\mu} = \pm 1$ with equal probability. We introduce the local stability parameter as

$$\Delta^{\mu} = h^{\mu}\xi^{\mu} \tag{2.3}$$

where h^{μ} is given by (2.1b) with S_k^{μ}. If the neuron generates ξ^{μ} in response to $\boldsymbol{S}^{\mu}$, i.e., $\Delta^{\mu} > 0$, then we say that the μth pattern is stored. If Δ^{μ} is a large positive number then high stability of storage against changes in either the couplings or the inputs can be assumed. Large stability is associated with large basin of attraction in memory networks.[5] Given the ensemble of patterns, the local stability parameter obeys some distribution $\rho(\Delta)$.[25] If the number of patterns M is of order N then it is useful to introduce the relative number of examples

$$\alpha = M/N \tag{2.4}$$

called also load parameter. Since an overall positive factor of $\boldsymbol{J}$ does not change the output, we set the norm of $\boldsymbol{J}$ to $\sqrt{N}$, expressed by the prior distribution

$$w(\boldsymbol{J}) = C_N\, \delta(N - |\boldsymbol{J}|^2) \tag{2.5}$$

This is called spherical constraint. The factor C_N normalizes $w(\boldsymbol{J})$ to unity, it has no thermodynamical significance besides setting the zero point of the entropy scale. Given the distribution of patterns and the length of $\boldsymbol{J}$ it can be easily seen that the normalization in (2.1b) results in h values of typically $O(1)$.

Storage with minimal error can be formulated as an optimization task by our introducing an error measure. If we treat all patterns in the same way we obtain what is called the equilibrium problem. The associated Hamiltonian, or cost function, is

$$\mathscr{H} = \sum_{\mu=1}^{M} V(\Delta^{\mu}) \tag{2.6}$$

where the potential $V(\Delta^{\mu})$ gives the error on a single pattern $\boldsymbol{S}^{\mu}, \xi^{\mu}$. Obviously, $V(y) = 0$ for $y > 0$ in the original storage problem. One can also impose a bound κ for the local stability, i.e., $V(y)$ is set to be zero for

$y > \kappa$, if $\kappa > 0$ this is obviously stricter than the original storage criterion. Generically, $V(y)$ should be monotonically decreasing for $y < \kappa$. In this paper specific results will be presented on the error counting, or Gardner–Derrida,[4, 12] potential

$$V(y) = \theta(\kappa - y) \tag{2.7}$$

where $\theta(y)$ is the Heaviside function. Given the potential (2.7) the aim is to minimize the number of patterns whose stability is below the bound κ. In the theoretical framework we shall keep the general form $V(y)$.

The partition function of the optimization task is

$$Z = \int d^N \boldsymbol{J}\, w(\boldsymbol{J}) \exp\left(-\beta \sum_{\mu=1}^{M} V(\Delta^\mu) \right) \tag{2.8}$$

with $\beta = 1/T$. Quenched average, $\langle \cdots \rangle_{\mathrm{qu}}$, is defined as the mean over the patterns. We deal with the idealized equilibrium of the system, when for large N the free energy, the energy, and the entropy are assumed to approach their quenched average. This property of self-averaging was proved rigorously only in special cases, see for example ref. 26, but it is the widely used basis in studies of the equilibrium thermodynamics of disordered systems.[6]

The replica method[6, 7] consists in writing the mean free energy per coupling as

$$f = -\lim_{N\to\infty} \frac{\langle \ln Z \rangle_{\mathrm{qu}}}{N\beta} = \lim_{N\to\infty} \lim_{n\to 0} \frac{1 - \langle Z^n \rangle_{\mathrm{qu}}}{nN\beta} \tag{2.9}$$

Denoting the thermal average with the Boltzmann weight in (2.8) by $\langle \cdots \rangle_{\mathrm{th}}$ the mean error per pattern can be written as

$$\varepsilon = \langle \langle V(\Delta) \rangle_{\mathrm{th}} \rangle_{\mathrm{qu}} = \frac{1}{\alpha} \frac{\partial \beta f}{\partial \beta} \tag{2.10}$$

The entropy is

$$s = \beta(\alpha\varepsilon - f) \tag{2.11}$$

For $T = 0$ and $\varepsilon = 0$ the volume of version space, i.e., the space of couplings that perfectly reproduce the examples is obtained as $\Omega = \exp(Ns)$. In general, Ns has the usual meaning of the logarithm of the volume with given error ε.

Introducing the overlap matrix $\mathbf{Q}$ of synaptic vectors

$$[\mathbf{Q}]_{ab} \equiv q_{ab} = \frac{1}{N} \sum_{k=1}^{N} J_{ak} J_{bk} \tag{2.12}$$

where the replica indices a, b go from 1 to n, we can express the free energy as the result of the minimization of the replica free energy[25, 22, 24]

$$f = \lim_{n \to 0} \frac{1}{n} \min_{\mathbf{Q}} f(\mathbf{Q}) \tag{2.13a}$$

$$f(\mathbf{Q}) = f_s(\mathbf{Q}) + \alpha f_e(\mathbf{Q}) \tag{2.13b}$$

$$f_s(\mathbf{Q}) = -(2\beta)^{-1} \ln \det \mathbf{Q} \tag{2.13c}$$

$$f_e(\mathbf{Q}) = -\frac{1}{\beta} \ln \int \frac{d^n x \, d^n y}{(2\pi)^n} \exp\left(-\beta \sum_{a=1}^{n} V(y_a) + i\boldsymbol{x}\boldsymbol{y} - \frac{1}{2} \boldsymbol{x}\mathbf{Q}\boldsymbol{x} \right) \tag{2.13d}$$

The subscripts s and e stand for entropic and energy-like terms. The entropic contribution (2.13c) arises because of the spherical constraint and the definition (2.12), it is indeed independent of the potential, while (2.13d) depends on it.

A central role is played by the probability density for the local stabilities[25, 22]

$$\rho(\Delta) = \langle\!\langle \delta(\Delta - h^1 \xi^1) \rangle_{\text{th}} \rangle_{\text{qu}} \tag{2.14}$$

where (2.1b) is understood. Due to the symmetry of the Hamiltonian with respect to the permutation of patterns we could choose $\mu = 1$ for convenience. It will turn out to be useful to interpret the integrand in (2.13d) as an effective Boltzmann weight and denote the average over this measure as

$$\langle\!\langle \cdots \rangle\!\rangle \tag{2.15}$$

where the $n \to 0$ limit is implied. A straightforward replica calculation shows (see ref. 27 for a pedagogic presentation) that the local stability distribution can be rewritten as

$$\rho(\Delta) = \langle\!\langle \delta(\Delta - y_1) \rangle\!\rangle \tag{2.16}$$

Here the subscript could be any replica index, for convenience we chose 1. Comparison with (2.14) allows an intuitive interpretation for the replica average $\langle\!\langle \cdots \rangle\!\rangle$, namely, this corresponds to the combined average over

thermal and quenched fluctuations. From (2.14) and (2.16) it follows that the combined thermal and quenched average in (2.10) boils down to

$$\varepsilon = \langle\langle V(y_1) \rangle\rangle = \int dy\, \rho(y)\, V(y) \tag{2.17}$$

The free energy (2.13) was calculated within the replica symmetric ansatz for the error counting potential (2.7) and the capacity, i.e., the maximal α with $\varepsilon = 0$ at $T = 0$ was determined in refs. 4 and 12. It has been shown that beyond capacity the replica symmetric solution is thermodynamically unstable.[28] One[22, 23] and two[19] step replica symmetry breaking solutions were presented, while ref. 19 proved that no finite step symmetry breaking ansatz can possibly be thermodynamically stable. We presented in ref. 24 a variational free energy functional without derivation that incorporated continuous replica symmetry breaking, but gave concrete results only in the high temperature, large α limit. In what follows we provide some background to the general theory and in the end properties of the ground state ($T = 0$) beyond capacity will also be described.

3. THE PARISI SCHEME

In this section we show how to evaluate the replica free energy $f_e(\mathsf{Q})$ of (2.13d) within Parisi's ansatz. The R step replica symmetry breaking form is[29]

$$\mathsf{Q} = \sum_{r=0}^{R+1} (q_r - q_{r-1})\, \mathsf{U}_{m_r} \otimes \mathsf{I}_{n/m_r} \tag{3.1}$$

where k subscripts k-dimensional matrices, I_k is the identity operator, all elements of U_k equal 1, $\otimes$ marks the direct product, and

$$q_{-1} = 0 \leqslant q_0 \leqslant q_1 \cdots \leqslant q_R \leqslant q_{R+1} = 1 \tag{3.2a}$$

$$m_{R+1} = 1 \leqslant m_R \leqslant m_{R-1} \cdots \leqslant m_1 \leqslant m_0 = n \tag{3.2b}$$

where the integer m_r is a divisor of m_{r-1}. The $n \to 0$ limit can be performed smoothly if instead of m_r we use $x_r = (n - m_r)/(n-1)$ for parametrization. Thus for arbitrary $n > 0$ we have the ordering

$$x_{R+1} = 1 \geqslant x_R \geqslant x_{R+1} \cdots \geqslant x_1 \geqslant x_0 = 0 \tag{3.3}$$

We consider the x_rs fixed along the $n \to 0$ limiting process, whence follows the formal n-dependence of the $m_r(n)$s, and for $n = 0$ we get $x_r = m_r(0)$. The inspection of the first few $R = 0, 1, 2$ cases[4, 22, 23, 19] allows, in the spirit of

Parisi's,[29] the generalization of the energy term (2.13d) in the replica free energy to arbitrary R as

$$f_e(\boldsymbol{q}, \boldsymbol{x}) = \lim_{n \to 0} \frac{1}{n} f_e(\mathbb{Q})$$

$$= -\frac{1}{\beta x_1} \int Dz_0 \ln \int Dz_1 \left[\int Dz_2 \cdots \left[\int Dz_{R+1} \right.\right.$$

$$\left.\left. \times \exp\left\{ -\beta V\left(\sum_{r=0}^{R+1} z_r \sqrt{q_r - q_{r-1}} \right) \right\} \right]^{x_R/x_{R+1}} \cdots \right]^{x_1/x_2} \tag{3.4}$$

where

$$Dz = \frac{dz\, e^{-1/2\, z^2}}{\sqrt{2\pi}} \tag{3.5}$$

This is the analog of Parisi's formula for the Sherrington–Kirkpatrick model, Eq. (11) in ref. 29; a comprehensive derivation will be presented elsewhere.[27] The energy term (2.13d) has become a function of the parameters in (3.2a) and (3.3). The evaluation of (3.4) can be done by iteration,

$$\psi_{r-1}(y) = \int Dz\, \psi_r(y + z \sqrt{q_r - q_{r-1}})^{x_r/x_{r+1}} \tag{3.6a}$$

$$\psi_{R+1}(y) = e^{-\beta V(y)} \tag{3.6b}$$

where $x_{R+2} = 1$ is understood. Then the sought free energy term is obtained as

$$f_e(\boldsymbol{q}, \boldsymbol{x}) = -\frac{1}{\beta x_1} \int Dz \ln \psi_0(z \sqrt{q_0}) \tag{3.7}$$

where $\boldsymbol{q} = (q_0, ..., q_R)$ and $\boldsymbol{x} = (x_1, ..., x_R)$.

The above iteration can be redressed as a partial differential equation. Parisi's order parameter function $x(q)$ is a concatenation of the $\boldsymbol{q}$ and $\boldsymbol{x}$ as

$$x(q) = \sum_{r=0}^{R} (x_{r+1} - x_r)\, \theta(q - q_r) \tag{3.8}$$

where $x_{-1} = 0$. Next we introduce $\psi(q, y)$ such that at the discontinuities

$$\psi(q_r^{+0}, y) = \psi_r(y) \tag{3.9a}$$

$$\psi(q_r^{-0}, y) = \psi(q_r^{+0}, y)^{x(q_r^{-0})/x(q_r^{+0})} \tag{3.9b}$$

that is, $\psi(q, y)^{1/x(q)}$ is continuous in q. Furthermore, along the plateau in the open interval (q_{r-1}, q_r)

$$\psi(q, y) = \int Dz\, \psi(q_r^{-0}, y + z\sqrt{q_r - q}) \tag{3.10}$$

It is easy to show that the field so defined satisfies the partial differential equation (PDE)

$$\partial_q \psi(q, y) = -\frac{1}{2}\partial_y^2 \psi(q, y) + \frac{\dot{x}(q)}{x(q)}\psi(q, y)\ln\psi(q, y) \tag{3.11a}$$

$$\psi(1, y) = e^{-\beta V(y)} \tag{3.11b}$$

which evolves from $q = 1$ to $q = 0$. Indeed, along the plateaus $\dot{x}(q) = 0$ when only the first term on the r.h.s. of (3.11a) remains, thus producing (3.10). Near jumps of $x(q)$ the second term dominates, and at a fixed y the resulting ordinary differential equation in the variable q is separable. Hence it follows that $\psi(q, y)^{1/x(q)}$ is continuous in q, thus (3.9b) is recovered. An equivalent field can be defined by

$$f(q, y) = -\frac{\ln\psi(q, y)}{\beta x(q)} \tag{3.12}$$

satisfying for a continuous potential $V(y)$ the PDE

$$\partial_q f(q, y) = -\tfrac{1}{2}\partial_y^2 f(q, y) + \tfrac{1}{2}\beta x(q)(\partial_y f(q, y))^2 \tag{3.13a}$$

$$f(1, y) = V(y) \tag{3.13b}$$

The fact that f denotes the free energy (2.13a) as well as the field $f(q, y)$ should not cause misunderstandings. Equation (3.13a) with initial condition $\ln 2\cosh\beta y$ has been discovered by Parisi[29] while studying the Sherrington–Kirkpatrick model.

If the potential $V(y)$ is not continuous, the PDE (3.13) holds only from any $q^* < 1$ onward where $\ln\psi(q^*, y)$ is continuous in y. In the generic case such is q_R, so the evolution along the first plateau from 1 to q_R, where $x(q) \equiv 1$, is to be done explicitly as

$$f(q_R, y) = -\frac{\ln\psi(q_R, y)}{\beta x_{R+1}} = -\beta^{-1}\ln\int Dz\, e^{-\beta V(y + z\sqrt{1 - q_R})} \tag{3.14}$$

and for $q < q_R$ the PDE (3.13a) with the initial condition (3.14) can be used. As a further alternative, one can introduce the field $m(q, y)$ as

$$\psi(q, y)\, m(q, y) = -\frac{\partial_y \psi(q, y)}{\beta x(q)} \tag{3.15}$$

which equals $\partial_y f(q, y)$ when continuity in y holds. Then the PDE (3.13a) should be replaced by

$$\partial_q f(q, y) = -\tfrac{1}{2}\, \partial_y m(q, y) + \tfrac{1}{2} \beta x m^2(q, y) \tag{3.16}$$

while the initial condition (3.13b) can be kept.

Finally we obtain the sought free energy term (2.13d) as a functional of the order parameter function

$$f_e[x(q)] = \lim_{n \to 0} \frac{1}{n} f_e(\mathcal{Q}) = f(0, 0) \tag{3.17}$$

It should be emphasized that the above PDEs do not require infinite refining of the partition by q_rs of the interval $(0, 1)$. They are valid for discrete as well as continuous replica symmetry breaking schemes, i.e., they admit $x(q)$ with steps and plateaus, as well as strictly monotonically increasing continuous segments.

The entropic term (2.13c) can also be cast in the form of the energy term (2.13d) with the substitution

$$e^{-\beta V(y)} = \sqrt{2\pi}\, \delta(y) \tag{3.18}$$

If we conceive the Dirac delta as a Gaussian with small variance, the initial condition (3.13b) becomes a quadratic function and the PDE (3.13a) can be solved analytically. The analogue of (3.17) for the entropic term, after going with the variance to zero, can be cast into

$$f_s[x(q)] = \lim_{n \to 0} \frac{1}{n} f_s(\mathcal{Q}) = -\frac{1}{2\beta} \int_0^1 dq \left[\frac{1}{D(q)} - \frac{1}{1-q} \right] \tag{3.19}$$

where

$$D(q) = \int_q^1 d\bar{q}\, x(\bar{q}) \tag{3.20}$$

The free energy of the neuron is then obtained as

$$f = \max_{x(q)} f[x(q)] \tag{3.21}$$

where the free energy functional is

$$f[x(q)] = f_s[x(q)] + \alpha f_e[x(q)] \tag{3.22}$$

with f_s and f_e defined in Eqs. (3.19) and (3.17). It is due to the $n \to 0$ limit that the maximization in (3.21) replaces the minimization by the matrix elements of Q in (2.13a), see, e.g., ref. 6.

4. LINEAR RESPONSE THEORY, STATIONARITY CONDITIONS, AND EXPECTATION VALUES

The least obvious part of the extremization condition (3.21) is the variation of $f(0, 0)$ by $x(q)$. This can be calculated from linear response theory for the PDE (3.13a). Moreover, linear response theory yields a technique to calculate replica averages as introduced in (2.15), essential for the evaluation of physical quantities.

The Green function for the PDE (3.13a) can be introduced formally as

$$\mathcal{G}(q_1, y_1; q_2, y_2) = \frac{\delta f(q_1, y_1)}{\delta f(q_2, y_2)} \tag{4.1}$$

whence $\mathcal{G}(q_1, y_1; q_2, y_2) = 0$ for $q_1 > q_2$. The Green function for the Parisi solution of the Sherrington–Kirkpatrick model has been studied in refs. 30 and 31. In the fore and hind variable pairs the Green function $\mathcal{G}(q_1, y_1; q_2, y_2)$ satisfies the respective PDEs

$$\partial_{q_1} \mathcal{G} = -\tfrac{1}{2} \partial_{y_1}^2 \mathcal{G} + \beta x(q_1)\, m(q_1, y_1)\, \partial_{y_1} \mathcal{G} - \delta(q_1 - q_2)\, \delta(y_1 - y_2) \tag{4.2a}$$

$$\partial_{q_2} \mathcal{G} = \tfrac{1}{2} \partial_{y_2}^2 \mathcal{G} + \beta x(q_2)\, \partial_{y_2}[m(q_2, y_2)\, \mathcal{G}] + \delta(q_1 - q_2)\, \delta(y_1 - y_2) \tag{4.2b}$$

where m is given in (3.15). The first equation without the Dirac delta excitation is the linearization of the PDE (3.13a). The minus sign of the Dirac deltas follows from the fact that (4.2a) evolves towards decreasing "time" q. The homogeneous part of (4.2b) is obtained from the requirement that

$$\mathcal{G}(q_1, y_1; q_3, y_3) = \int dy_2\, \mathcal{G}(q_1, y_1; q_2, y_2)\, \mathcal{G}(q_2, y_2; q_3, y_3) \tag{4.3}$$

does not depend on q_2, and the plus sign of the inhomogeneous term is due to the fact that evolution goes towards increasing q. The homogeneous parts of the two PDEs (4.2b) and (4.2b) are called adjoint to each other.

For the sake of simplicity we gave formula (4.1) for the case of continuous potential $V(y)$. If the potential is discontinuous then the definition (4.1) should and can be appropriately modified when a q-argument is near 1, but the PDEs (4.2) for the Green functions hold as they are.

The significance of the Green function is in that it helps to solve the linear PDE with the source term $h(q, y)$

$$\partial_q \vartheta(q, y) = -\tfrac{1}{2}\partial_y^2 \vartheta(q, y) + \beta x(q)\, m(q, y)\, \partial_y \vartheta(q, y) + h(q, y) \tag{4.4}$$

as

$$\vartheta(q, y) = \int dy_1 \mathscr{G}_\varphi(q, y; 1, y_1)\, \vartheta(1, y_1) - \int_q^1 dq_1 \int dy_1 \mathscr{G}_\varphi(q, y; q_1, y_1)\, h(q_1, y_1) \tag{4.5}$$

A prominent role will be played by

$$P(q, y) = \mathscr{G}(0, 0; q, y) \tag{4.6}$$

which solves the PDE (4.2b) with $q_1 = y_1 = 0$, i.e., with initial condition

$$P(0, y) = \delta(y) \tag{4.7}$$

This function first appeared in the context of the Sherrington–Kirkpatrick model in ref. 32. Note that the PDE for $P(q, y)$ is in fact a Fokker–Planck equation, producing a nonnegative solution and conserving the norm $\int dy\, P(q, y) \equiv 1$ for all qs. This suggests the intuitive interpretation of $P(q, y)$ as probability density of y.

Now we are in the position to calculate the variation of the free energy functional (3.22). As to the energy term (3.17), by varying the functions $f(q, y)$ and $x(q)$ in the PDE (3.13a) one obtains (4.4) with $\vartheta = \delta f$, $\vartheta(1, y) = 0$, and $h = \frac{1}{2}\beta(\partial_y f)^2\, \delta x$. Hence (4.5) gives at $q = 0$, $y = 0$ the sought $\delta f(0, 0)/\delta x$. The variation of $f_s[x(q)]$ can be calculated straightforwardly, and, with the notation (4.6), we arrive at

$$F(q, [x(q)]) = \frac{2}{\beta}\frac{\delta f[x(q)]}{\delta x(q)} = \int_0^q \frac{d\bar{q}}{\beta^2 D(\bar{q})^2} - \alpha \int dy\, P(q, y)\, m(q, y)^2 \tag{4.8}$$

The stationarity condition in case $x(q)$ can be freely varied is thus

$$F(q, [x(q)]) = 0 \tag{4.9}$$

If in an interval I the $x(q)$ is supposed to have a plateau, we differentiate by the plateau value of $x(q)$ to get

$$\int_I dx\, F(q, [x(q)]) = 0 \tag{4.10}$$

In isolated points q_r where plateaus meet (4.9) should hold pointwise. This summarizes the stationarity conditions for an arbitrary order parameter function, i.e., arbitrary replica symmetry broken scheme of Parisi type.

We have seen in Section 2 instances when the double, thermal and quenched, average could be replaced by the replica average (2.15). Replica averages can be calculated by the Green function technique as described below. The procedure can be viewed as the generalization of the ground-breaking results from refs. 30 and 33, where the local magnetization and some of its moments in the Parisi phase of the Sherrington–Kirkpatrick model were evaluated.

A simple case is when $\langle\langle A(y_a)\rangle\rangle$ is to be calculated for an arbitrary function $A(y)$. Because of the symmetry with respect to the permutation of single replica indices we have $\langle\langle A(y_a)\rangle\rangle = \langle\langle n^{-1}\sum_{a=1}^{n} A(y_a)\rangle\rangle$. This quantity can be easily evaluated if one replaces in (2.13d) $V(y)$ by $V(y)+\lambda A(y)$, thus obtains $f_e(\mathsf{Q};\lambda)$, and calculates its initial slope at $\lambda = 0$. Reversing the limits $n \to 0$ and $\lambda \to 0$ then using the first equality of (3.4) and Eq. (3.17) we get

$$\begin{aligned}\langle\langle A(y_a)\rangle\rangle &= \lim_{n\to 0} \frac{1}{n} \left.\frac{\partial f_e(\mathsf{Q};\lambda)}{\partial \lambda}\right|_{\lambda=0} = \left.\frac{\partial f(0,0;\lambda)}{\partial \lambda}\right|_{\lambda=0} \\ &= \int dy\, \frac{\delta f(0,0)}{\delta f(1,y)} A(y) = \int dy\, P(1,y)\, A(y)\end{aligned} \tag{4.11}$$

The third equality comes from the fact that λ is in the initial condition for $f(q, y)$ at $q = 1$, and the last one comes from the definitions for the Green function (4.1) and for $P(q, y)$ (4.6). Again, for the sake of brevity we gave the derivation for continuous potential $V(y)$, however, the result holds also for discontinuous ones. Immediately follows from (2.16) the formula for the probability density of the local stabilities

$$\rho(y) = P(1, y) \tag{4.12}$$

and thus from (2.17)

$$\varepsilon = \int dy\, P(1, y)\, V(y) \tag{4.13}$$

From the practical viewpoint interesting are effective averages of products like

$$\langle\!\langle A_1(y_{a_1})\, A_2(y_{a_2}) \cdots A_k(y_{a_k}) \rangle\!\rangle \tag{4.14}$$

One can show by using elementary properties of the Fourier transformation that the average of a product of x_as over the effective Boltzmann weight on the r.h.s. of (2.13d) can be expressed as averages over functions of variables y_as. Thus the knowledge how to evaluate (4.14) also resolves the problem of averages of polynomials in x_as. The latter quantities are of importance because they appear when the replica free energy is differentiated in terms of q_{ab}s.

Here we shall only describe the recipe for calculating (4.14), details can be found in ref. 27. If $k=2$ then the average depends on $q=q_{a_1a_2}$ and is given by the formula

$$\begin{aligned} C_{12}(q) &= \langle\!\langle A_1(y_{a_1})\, A_2(y_{a_2}) \rangle\!\rangle \\ &= \int dy\, dy_2\, dy_3\, P(q,\, y)\, \mathcal{G}(q,\, y;\, 1,\, y_1)\, \mathcal{G}(q,\, y;\, 1,\, y_2)\, A_1(y_1)\, A_2(y_2) \end{aligned} \tag{4.15}$$

Of such type is $\partial f_e(\mathsf{Q})/\partial q_{a_1a_2}$ where $A_1(y)=A_2(y)=i\cdot m(1,\, y)$, see (3.15) for definition, whence we obtain the second term in $F(q,\, [x(q)])$ given in Eq. (4.8). This is related to the fact that the stationarity condition can also be obtained by first differentiating the replica free energy (2.13b) by q_{ab} and then equating the result to zero. For $k=3$ suppose without restricting generality that

$$q=q_{a_1a_3}=q_{a_2a_3}<\bar{q}=q_{a_1a_2} \tag{4.16}$$

Then we have

$$\begin{aligned} C_{123}(q,\, \bar{q}) &= \langle\!\langle A_1(y_{a_1})\, A_2(y_{a_2})\, A_3(y_{a_3}) \rangle\!\rangle \\ &= \int dy\, d\bar{y}\, dy_1\, dy_2\, dy_3\, P(q,\, y)\, \mathcal{G}(q,\, y;\, 1,\, y_3) \\ &\quad \times \mathcal{G}(q,\, y;\, \bar{q},\, \bar{y})\, \mathcal{G}(\bar{q},\, \bar{y};\, 1,\, y_1)\, \mathcal{G}(\bar{q},\, \bar{y};\, 1,\, y_2) \\ &\quad \times A_1(y_1)\, A_2(y_2)\, A_3(y_3) \end{aligned} \tag{4.17}$$

Special versions of the above formulas, for the case of the second and third moments of the magnetization in the Sherrington–Kirkpatrick model, were worked out in refs. 30 and 33. The integrals in (4.17) admit a simple graphic representation as shown on Fig. 1.

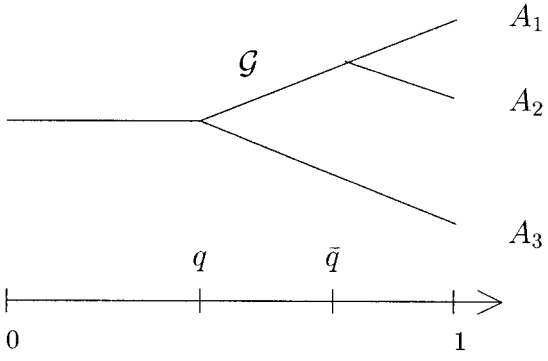

Fig. 1. Correlation function $C_{123}(q, \bar{q})$. The right endpoints correspond to the functions to be averaged and lines are Green functions (only one of them is labeled by $\mathcal{G}$ in the figure). All nodes, including the right endpoints, have ys which should be integrated over. The leftmost line is the function $P(q, y)$, or equivalently, this line is also a Green function and a $\delta(y)$ is associated with the leftmost endpoint.

The cases considered above are the effective average for $k=1$, given by Eq. (4.11), represented by one line, and for $k=2$, as calculated in (4.15), represented by a fork with one handle and two branches. Averages of more than three functions can be analogously constructed, for a given k a graph has $k+1$ "legs." Obviously there are two topologically possible graphs for $k=4$, depending on the overlaps $q_{a_i a_j}$, and more for larger ks.

The ability to calculate $k=4$ effective averages allows us to study linear stability of the replica free energy (2.13b) at the stationary Q. Using the results of ref. 34 on ultrametric matrices we expressed the so called replicon eigenvalues in terms of Green functions. While a general proof of the fact that there are no negative eigenvalues in the case of continuous replica symmetry breaking, i.e., when $x(q)$ has a continuously increasing segment, is not available, in the high temperature limit[24, 27] we confirmed the absence of linear instability against replicons whenever we encountered such a stationary state. For any temperatures we recovered analytically the zero eigenvalues, corresponding to Goldstone modes, as well as the lowest order Ward–Takahashi identities predicted by algebra.[35]

The generalization of (4.14) to non-factorizable functions is straightforward. For example, such functions would simply replace the products of A_k's in (4.15) and (4.17).

5. LOW TEMPERATURE RESULTS

In ref. 24 we have shown that in the high temperature limit, i.e., for $\alpha, \beta \to \infty$ with $\gamma = \alpha\beta^2$ finite, the problem simplifies to the extent that if $x(q)$

has a continuously increasing segment, this can be given in a closed analytic form. In that limit the problem becomes equivalent to the spherical, multi-p-spin interaction spin glass,[36] where a similar observation has been made. Four different phases have been found[24] for the error counting potential (2.7): for small γ replica symmetry holds, and for $|\kappa| < 2$ and large γ there is a Parisi phase with a single continuously increasing segment of $x(q)$ between the trivial plateaus $x \equiv 0$ and $x \equiv 1$. When $|\kappa| > 2$ there is also a narrow one-step replica symmetry broken regime, and for large enough γs equilibrium is characterized by an $x(q)$ that is a concatenation of a nontrivial plateau and a continuously increasing segment. While for $T > 0$ there cannot be error free storage, it is plausible to conceive the replica symmetric regime as the continuation of the $T = 0$ phase of perfect storage and the symmetry broken phases as the analog of the regime at $T = 0$ beyond capacity.

In the case of a $V(y)$ potential that vanishes for $y > \kappa$ the limit of capacity is given for $\kappa \geqslant 0$ at $T = 0$ by[4]

$$\alpha_c(\kappa) = \left(\int_{-\infty}^{\kappa} Dt(\kappa - t)^2 \right)^{-1} \tag{5.1}$$

as it follows from the replica symmetric solution when $q \to 1$. This formula also gives the limit of the de Almeida–Thouless (AT) local stability[28] in the case of the potential (2.7). For $\kappa = 0$ one has $\alpha_c = 2$, and, for increasing κ, $\alpha_c(\kappa)$ understandably decreases. As it has been already mentioned, beyond capacity none of the finite-step replica symmetry breaking schemes gives a locally stable equilibrium state.[19] Thus in this regime the order parameter function is no longer of the step-like form of (3.8), rather it has a continuously increasing part. This makes it necessary to numerically solve the extremization problem (3.21).

The ground state ($T = 0$) has its special scaling properties. The PDE (3.13a) stays meaningful if for $q < 1$ the function $\beta x(q)$ does not diverge, implying that $x(q)$ goes to zero. Given the meaning of $x(q)$ as the probability of a q being in the interval $(0, q)$,[6] we can say that at $T = 0$ the overlap q is 1 with probability 1 for all $\alpha > \alpha_c$. Nevertheless, a physically meaningful order parameter is obtained after scaling by β. Firstly we introduce for $T > 0$ the parameters of the classic Parisi shape as

$$\begin{aligned} &x(q) \equiv 0 && \text{if} \quad 0 \leqslant q < q_{(0)} \\ &\dot{x}(q) > 0 \text{ and } x_{(0)} < x(q) < x_{(1)} && \text{if} \quad q_{(0)} < q < q_{(1)} \\ &x(q) \equiv 1 && \text{if} \quad q_{(1)} < q \leqslant 1 \end{aligned} \tag{5.2}$$

The scaled quantities

$$q(t) = q_{(1)} - (q_{(1)} - q_{(0)})(1 - (1 + q_{(1)})\, t + q_{(1)} t^2), \qquad 0 \leqslant t \leqslant 1 \tag{5.3a}$$

$$\xi(t) = \beta x(q(t))\, \dot{q}(t) \tag{5.3b}$$

$$\eta = \beta(1 - q_{(1)}) \tag{5.3c}$$

$$\Delta(t) = \beta D(q(t)) = \int_t^1 \xi(\bar{t})\, d\bar{t} + \eta \tag{5.3d}$$

are expectedly regular even in the $T \to 0$ limit, when $q_{(1)} \to 1$. Note that there is an arbitrariness in the parametrization by t, the main features being that $q(0) = q_{(0)}$, $q(1) = q_{(1)}$, and $\dot{q}(1) = 0$. With this parametrization the PDE (3.13) becomes

$$\partial_t f(t, y) = -\frac{1}{2}\, \dot{q}(t)\, \partial_y^2 f(t, y) + \frac{1}{2}\, \xi(t)(\partial_y f(t, y))^2 \tag{5.4a}$$

$$f(1, y) = -\frac{1}{\beta} \ln \int Dz\, e^{-\beta V(y + z\sqrt{1 - q_{(1)}})} \tag{5.4b}$$

For $T = 0$ the initial condition becomes

$$f(t = 1, y)\,|_{T=0}$$

$$= \min_{\bar{y}} \left(V(\bar{y}) + \frac{(y - \bar{y})^2}{2\eta} \right) = \begin{cases} 1 & \text{if} \quad y \leqslant \kappa - \sqrt{2\eta} \\ \dfrac{(\kappa - y)^2}{2\eta} & \text{if} \quad \kappa - \sqrt{2\eta} \leqslant y \leqslant \kappa \\ 0 & \text{if} \quad y \geqslant \kappa \end{cases} \tag{5.5}$$

where (2.7) was substituted to get the second equality. By Gaussian integration hence the replica symmetric solution can be obtained.[4, 12, 25] Note that $f(t = 1, y)$ is a continuous function even though $V(y)$ is the step function. Although we used the same symbols for functions of q and t, misunderstanding are avoided by our marking which argument we mean.

The PDE for $P(q, y)$ follows from the definition (4.6) and from the evolution equation of the Green function (4.2b). The latter can be properly rescaled according to (5.3), yielding in principle the solution $P(t, y)$. The

probability density of local stabilities is obtained by evolving $P(t=1, y) = P(q=q_{(1)}, y)$ to $P(q=1, y)$. At $T=0$ with (2.7) we have

$$\rho(\Delta) = \begin{cases} P(1, \Delta) & \text{if} \quad \Delta \leqslant \kappa - \sqrt{2\eta} \\ 0 & \text{if} \quad \kappa - \sqrt{2\eta} < \Delta < \kappa \\ P(1, \Delta) + \delta(\Delta - \kappa) \int_{\kappa - \sqrt{2\eta}}^{\kappa} d\bar{y}\, P(1, \bar{y}) & \text{if} \quad \Delta \geqslant \kappa \end{cases} \tag{5.6}$$

where $P(1, \Delta) = P(t=1, \Delta)$ is understood. For arbitrarily small $T>0$ the gap in the support of $\rho(\Delta)$ immediately vanishes. For details we again refer to ref. 27.

The numerical extremization was done with complementing the free energy functional (3.22) by constraints. The PDE (5.4) was added giving rise to a Lagrange multiplier field. This field can be shown to be just $P(t, y)$, see refs. 24, 27, and 31. Further technical requirements are $x(q_{(0)}) \geqslant 0$, $x(q_{(1)}) \leqslant 1$, and $\dot{x}(q) \geqslant 0$, which were taken into account by soft

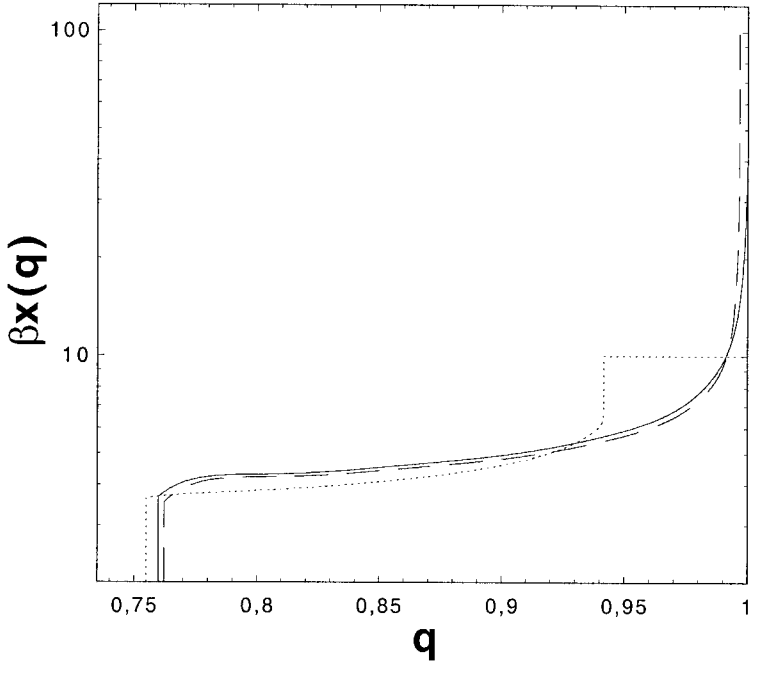

Fig. 2. Scaled order parameter function $x(q)$ for $\kappa=0$, $\alpha=3$ at $T=0$ (solid), $T=0.01$ (dashed), and $T=0.1$ (dotted).

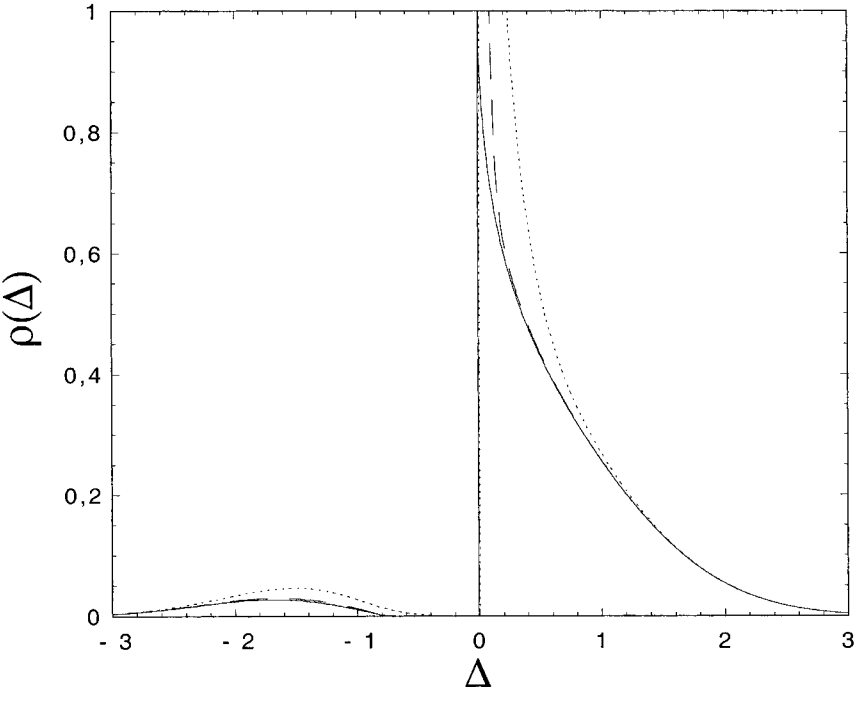

Fig. 3. Density of local stabilities $\rho(\Delta)$ from theory for $\kappa = 0$, $\alpha = 3$ at $T = 0$ (solid), $T = 0.01$ (dashed), and $T = 0.1$ (dotted).

constraints. A few results for the Gardner–Derrida potential (2.7), with $\kappa = 0$, $\alpha = 3$, are displayed on Figs. 2 and 3 for various low temperatures. Note that the point ($\kappa = 0$, $\alpha = 3$) lies beyond capacity, while beyond about $\beta^{-1} = T = 0.2$ the RS solution satisfies the AT stability condition.[27]

On Fig. 2 the scaled order parameter function $\beta x(q)$ is shown. For small temperatures the replica symmetric solution is AT unstable, and we indeed obtain the Parisi form (5.2) for $x(q)$. At $q_{(0)}$ near 0.75 the functions $x(q)$ jump to zero and remains there as q further decreases. The upper plateau with $x(q) \equiv 1$ starts at $q_{(1)}$. The consistency of the scaling (5.3) is confirmed by our finding at $T = 0$ finite values for $\beta x(q)$ if $q < 1$ and for $\eta \approx 0.26$. Interestingly, the curved segment of $\beta x(q)$ does not change much with increasing temperature, the main effect being the decrease of $q_{(1)}$.

The local stability distribution $\rho(\Delta)$ of (2.16) is displayed on Fig. 3. It was numerically obtained for $T = 0$ from (5.6) and for $T > 0$ from the original formula (4.12), with the same parameter values as in Fig. 3. For the sake of better visibility of the other details, the very high peaks near

$\Delta = \kappa = 0$ for $T = 0.1$ and $T = 0.01$ as well as the corresponding δ-peak for $T = 0$ have been omitted from this plot. For $|\Delta| > 3$ the curves approach zero very quickly. A true gap with $\rho(\Delta) = 0$ to the left of $\Delta = 0$ develops only if $T = 0$, while for the positive T-values $\rho(\Delta)$ is positive albeit small there. While the gap at $T = 0$ is present in $R = 0$, 1, 2 step replica symmetry breaking schemes, see refs. 25, 22, and 19 respectively, in all these cases a jump appears near the lower edge. This can be associated with the thermodynamic instability of those saddle points.[37] Our present solution gives linearly vanishing $\rho(\Delta)$ at the lower edge, signaling the absence of replicon instability.[27]

In order to compare theory with practice we performed a medium scale simulation. The standard Hebbian algorithm was modified by Wendemuth[38] to provide convergence for negative stabilities. Since we chose $\kappa = 1$, the final steps during stabilization of a pattern went on with $\Delta > 0$, so in our case the modification was not essential. The algorithm goes as follows. Firstly random patterns (2.2) are generated uniformly from an interval centered about zero and normalized as $\sum_{k=1}^{N} (S_k^\mu)^2 = N$, and all outputs ξ^μ are taken uniformly 1. This does not restrict generality since S_k^μ have random signs. The initial coupling vector is set to be proportional to

$$J_k(0) \propto \sum_{\mu=1}^{M} S_k^\mu \tag{5.7}$$

such that its Eucledian norm is N. In the tth step of the algorithm one calculates the local stabilities

$$\Delta^\mu(t) = \frac{\boldsymbol{J}(t) \cdot \boldsymbol{S}^\mu}{|\boldsymbol{J}(t)|} \tag{5.8}$$

and selects the least unstable pattern, i.e., the one with the largest $\Delta^\mu(t) < \kappa$. Let us denote its index by μ_0, whose argument t we omit. Next one augments the couplings as follows. If $\Delta^{\mu_0}(t) > 0$ then

$$\boldsymbol{J}(t+1) = \boldsymbol{J}(t) + \lambda \boldsymbol{S}^{\mu_0} \tag{5.9}$$

and if $\Delta^{\mu_0}(t) < 0$ we have following Wendemuth

$$\boldsymbol{J}(t+1) = \boldsymbol{J}(t) + \lambda \left(\boldsymbol{S}^{\mu_0} + \boldsymbol{J}(t) \frac{N/|\boldsymbol{J}(t)| - \Delta^{\mu_0}(t)}{|\boldsymbol{J}(t)| - \Delta^{\mu_0}(t)} \right) \tag{5.10}$$

Here λ is the gain parameter, the overall scale of increments of $\boldsymbol{J}$. In ref. 38 the gain parameter was $\lambda = N^{-3/2}$, after experimentation we chose $\lambda = N^{-1}$. Such an increase in the gain parameter did not endanger, rather sped up

convergence. At time $t+1$ we again look for the least unstable pattern, and so on. The algorithm goes on until it gets stuck with one pattern that we are not able to stabilize in a reasonable time. The intuitive idea behind the algorithm is that since an unstable pattern counts as error irrespective of the distance of the stability parameter Δ^μ from κ, one assumes that it is the easiest to stabilize the pattern with Δ^μ closest to κ. So one may hope that thus the largest possible number of patterns can be stabilized.

The program ran on 28 PCs in parallel, each having an AMD K6 processor of 333 MHz, during about one day. Figure 4 shows both the theoretical curve and the results of the simulation for $\alpha=\kappa=1$, a point known to fall beyond the capacity curve (5.1). The full line is the result of numerical extremization of the free energy functional in the way Fig. 3 was obtained. The Dirac delta peak of the theoretical probability density at κ is not illustrated. The discontinuous lines represent the histograms for the local stabilities from simulation for two sizes, $M=N=500$ and 1000, after normalization.

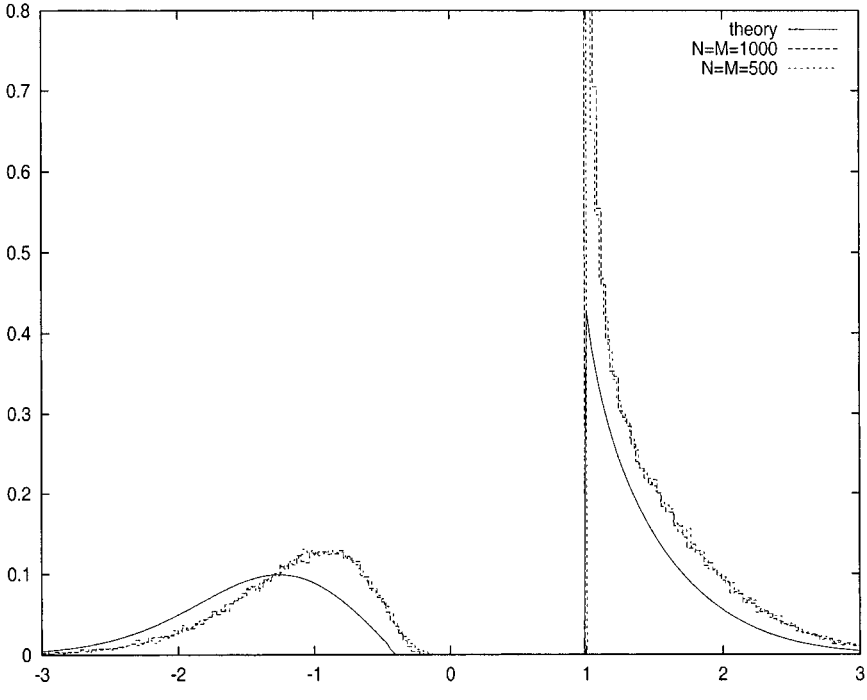

Fig. 4. Density of local stabilities $\rho(\Delta)$ at $\alpha=\kappa=1$, axes as in Fig. 3. The theoretical prediction is given by the full line. The two empirical densities are normalized histograms, taken with $M=N=500$ and 1000.

The closeness of the two histograms demonstrates that size effects were probably not the cause for the systematic difference between theory and the numerical experiment. A possible ground for the discrepancy is that the algorithm may have been halted prematurely. However, the time necessary for the stabilization of patterns was allowed to grow for each subsequent pattern, and the algorithm was ended only when stabilization did not occur even within the multiple of such an extrapolated time. Another possible reason for the deviation may be that the algorithm got stuck in a "local optimum" without being able to globally maximize the number of stable patterns. In this regard several modified initial conditions were tested but the number of stabilized patterns did not grow in the end. A source of concern can be that the built in random number generator of the C compiler was used; we did not test other routines for this purpose. As to the algorithm, despite its intuitive appeal, there is no proof that it would be able to globally minimize the Hamiltonian (2.6) with error measure (2.7). Furthermore, it is likely that with the present parameters the learning task is an NP-complete problem,[5, 38, 39] thus explaining imperfect convergence.

We emphasize that the results represent a significant improvement with respect to the earlier simulation in ref. 39. The error per example ε found in ref. 39 is about 0.21, while the present data correspond to 0.15 and theory predicts 0.1358. Thus the deviation between simulation and theory has been decreased by 80%. That means that we stabilized more patterns than ref. 39, although, given the difference between the theoretical and simulation results, we still could not find the global optimum. Furthermore, an important feature of the density $\rho(\Delta)$ is that it should continuously vanish, with a nonzero slope, at the lower edge of the gap. This property is reproduced by the simulation data, in a sharper fashion with the larger $M = N = 1000$ size, while the value of the edge remains slightly overestimated.

ACKNOWLEDGMENTS

The authors acknowledge support by OTKA Grant T017272 (G.Gy.) and from a special grant for young scientists from the University of Augsburg and by the State of Bavaria within the postgraduate scheme Graduiertenkolleg GRK283 "Nonlinear Problems in Analysis, Geometry, and Physics" (P.R.). It is a pleasure to thank F. Csikor and Z. Fodor for offering us for the simulation their PC farm, supported by Grants OTKA-T22929 and FKFP-0128/1997. Thanks are due to F. Pázmándi for his pointing out and discussing with us the problem of a discontinuity in the potential.

REFERENCES

1. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**:2554 (1982).
2. W. A. Little, *Math. Biosci.* **19**:101 (1974).
3. W. S. McCulloch and W. Pitts, *Bull. Mathem. Biophys.* **5**:115 (1943).
4. E. Gardner, *J. Phys. A* **21**:257 (1988).
5. J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison–Wesley, Reading, MA, 1991).
6. M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
7. K. H. Fischer and J. A. Hertz, *Spin Glasses* (Cambridge University Press, Cambridge, UK, 1991).
8. F. Rosenblatt, *Principles of Neurodynamics* (Spartan, New York, 1962).
9. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Nature* **323**:533 (1986).
10. T. L. H. Watkin, A. Rau, and M. Biehl, *Rev. Mod. Phys.* **65**:499 (1993).
11. S. Amari, *Neural Computation* **10**:251 (1998).
12. E. Gardner and B. Derrida, *J. Phys. A* **21**:271 (1988).
13. M. L. Minsky and S. A. Papert, *Perceptrons* (MIT Press, Cambridge, MA, 1969).
14. M. Biehl and N. Caticha, in *The Handbook of Brain Theory and Neural Networks*, 2nd edn., M. A. Arbib, ed. (MIT Press, Cambridge, MA, 1999).
15. D. Malzahn, A. Engel, and I. Kanter, *Phys. Rev. E* **55**:7369 (1997).
16. B. Schottky, *J. Phys. A* **28**:4515 (1995).
17. A. H. L. West and D. Saad, *J. Phys. A* **31**:8977 (1998).
18. O. Winther, B. Lautrup, and J.-B. Zhang, *Phys. Rev. E* **55**:836 (1997).
19. W. Whyte and D. Sherrington, *J. Phys. A* **29**:3063 (1996).
20. M. Opper and W. Kinzel, in *Models of Neural Networks*, Vol. III, E. Domany, J. L. van Hemmen, and K. Schulten, eds. (Springer-Verlag, New York, 1996).
21. G. Györgyi and N. Tishby, Statistical mechanics of learning a rule, in *Proceedings of the STATPHYS-17 Workshop on Neural Networks and Spin Glasses*, W. K. Theumann and R. Köberle, eds. (Singapore–Teaneck, NJ–Hongkong, 1990, World Scientific).
22. P. Majer, A. Engel, and A. Zippelius, *J. Phys. A* **26**:7405 (1993).
23. R. Erichsen and W. K. Theumann, *J. Phys. A* **26**:L61 (1993).
24. G. Györgyi and P. Reimann, *Phys. Rev. Lett.* **79**:2746 (1997).
25. M. Griniasty and H. Gutfreund, *J. Phys. A* **24**:715 (1991).
26. M. Aizenman, J. L. Lebowitz, and D. Ruelle, *Comm. Math. Phys.* **112**:3 (1987).
27. G. Györgyi, *Phys. Rep.*, to be published.
28. M. Bouten, *J. Phys. A* **27**:6021 (1994).
29. G. Parisi, *J. Phys. A* **13**:L115 (1980).
30. J. R. L. de Almeida and E. J. S. Lage, *J. Phys. C* **16**:939 (1983).
31. H.-J. Sommers and W. Dupont, *J. Phys. C* **17**:5785 (1984).
32. H. Sompolinsky and A. Zippelius, *Phys. Rev. Lett.* **47**:935 (1981).
33. M. Mézard and M. A. Virasoro, *J. Phys. France* **46**:1293 (1985).
34. T. Temesvári, C. De Dominicis, and I. Kondor, *J. Phys. A* **27**:7569 (1994).
35. C. De Dominicis, T. Temesvári, and I. Kondor, *J. Phys. IV France* **8**:Pr6-13 (1998).
36. Th. M. Nieuwenhuizen, *Phys. Rev. Lett.* **74**:4289 (1995).
37. A. Engel has pointed out this fact to us.
38. A. Wendemuth, *J. Phys. A* **28**:5423 (1995).
39. A. Wendemuth, *J. Phys. A* **28**:5485 (1995).